

# SCIENCE PAPER OR BIG DATA?

Assessing invasion dynamics using observational data

Charlotte Clubley, Louise Firth, Louisa Wood,  
David Bilton, Tiago Silva, Antony Knights



UNIVERSITY OF  
PLYMOUTH



Centre for Environment  
Fisheries & Aquaculture  
Science

# The problem of non-natives

Global environmental change provides novel opportunities for spread

TRANSPORT

Establishment is required for invasion progression

ESTABLISHMENT

Understanding patterns of secondary spread is a key objective

SPREAD

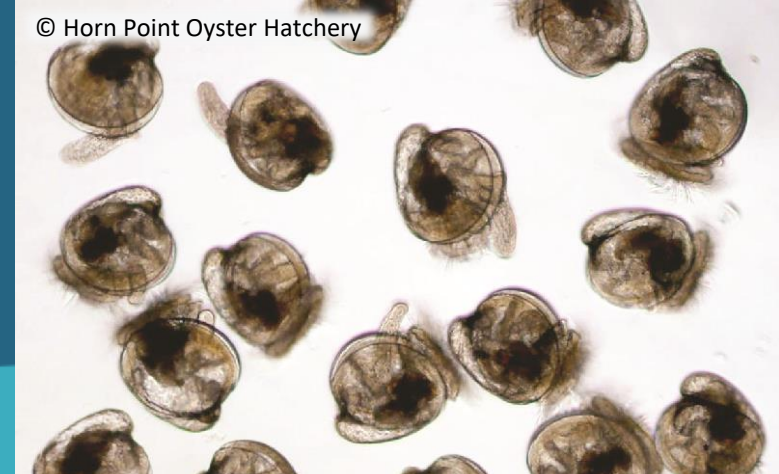


© Safety4Sea

© EcoShape



© Horn Point Oyster Hatchery



# The available data

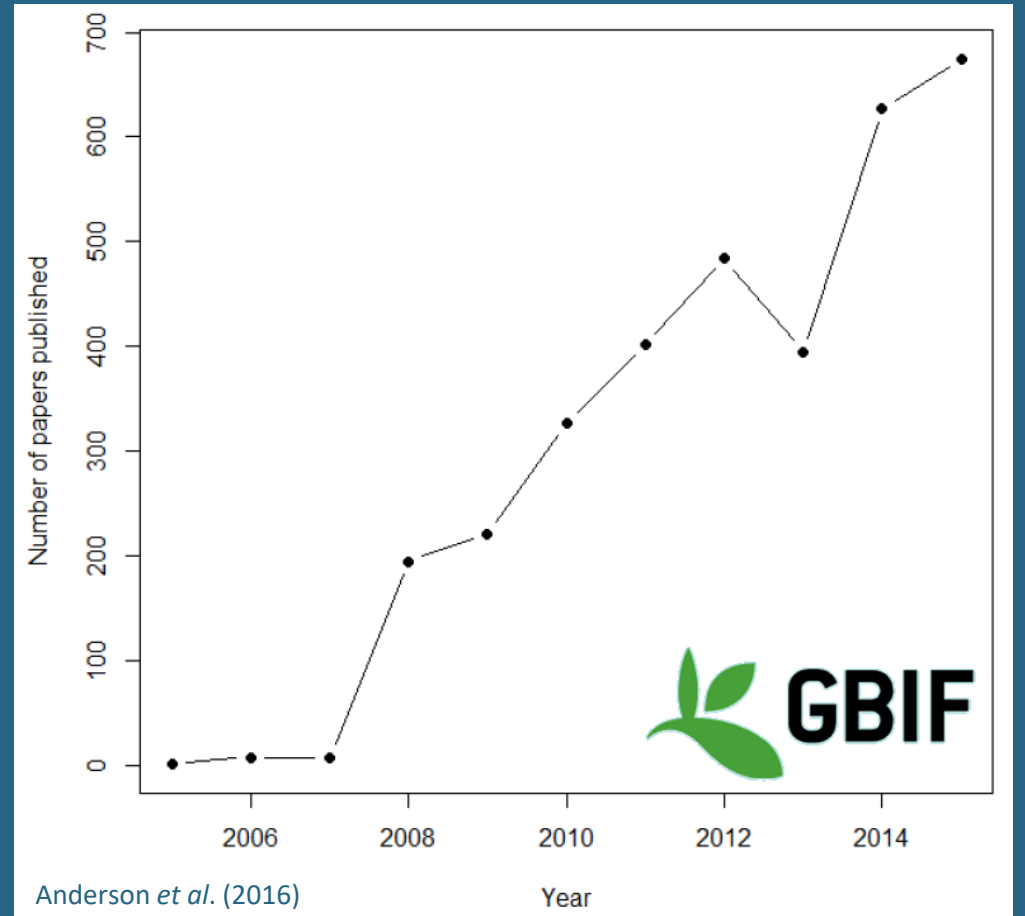
Estimates of spread require records of spatial distribution over time



Detailed introduction history of invasives is often available



Quantity and quality of data influence accuracy of estimates



Numerous scientists rely heavily on online data repositories

**(1) Identify discrepancies between data types**

Comparison of introduction history from peer reviewed literature ("Science Paper") and online databases ("Big Data")

**(2) Do discrepancies affect spread estimates?**

Comparison of spread estimates calculated using either data source



**A framework for comparison of data sources**

# Pacific oyster

## *Magallana (Crassostrea) gigas*

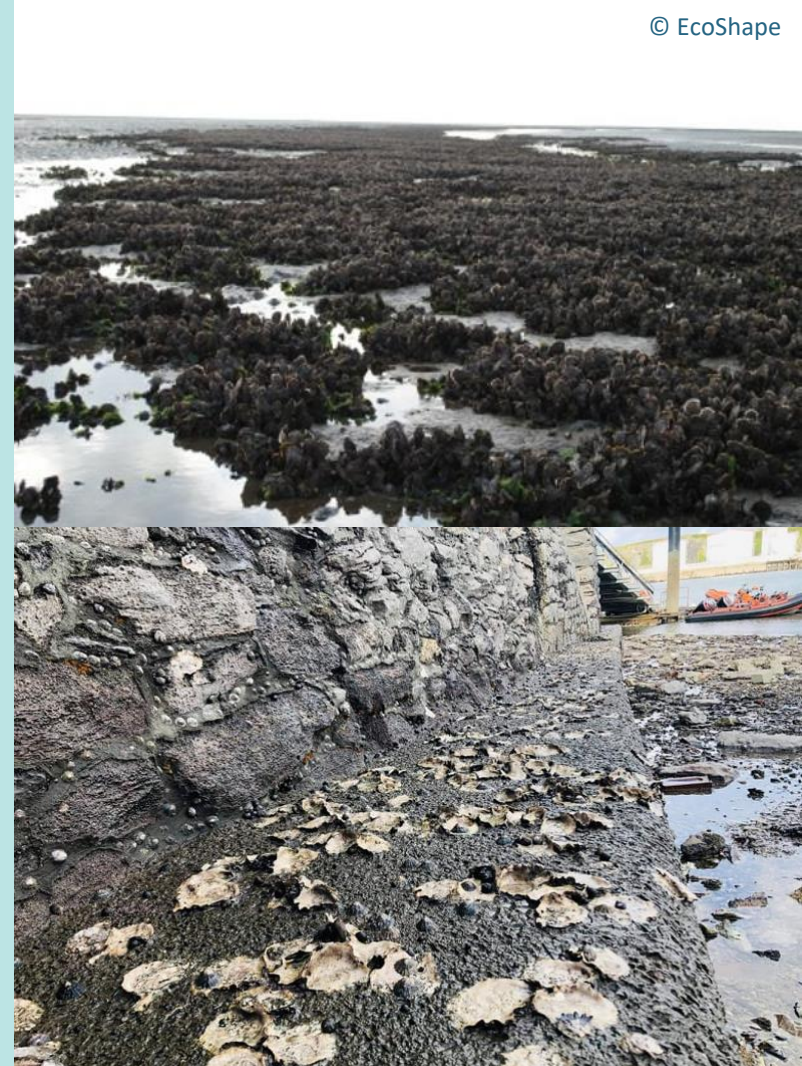
**‘One of the 100 worst alien species in Europe’**

*DAISIE inventory of Alien Invasive  
Species in Europe*

Introduced for aquaculture in the 1960s

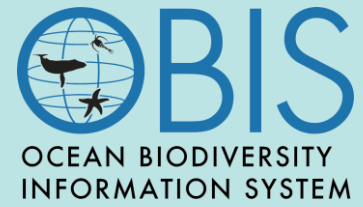
Potential impacts: biodiversity, economics,  
socio-economics

Eradication not likely or feasible





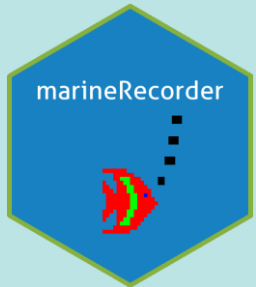
19,909  
records



3,760  
records



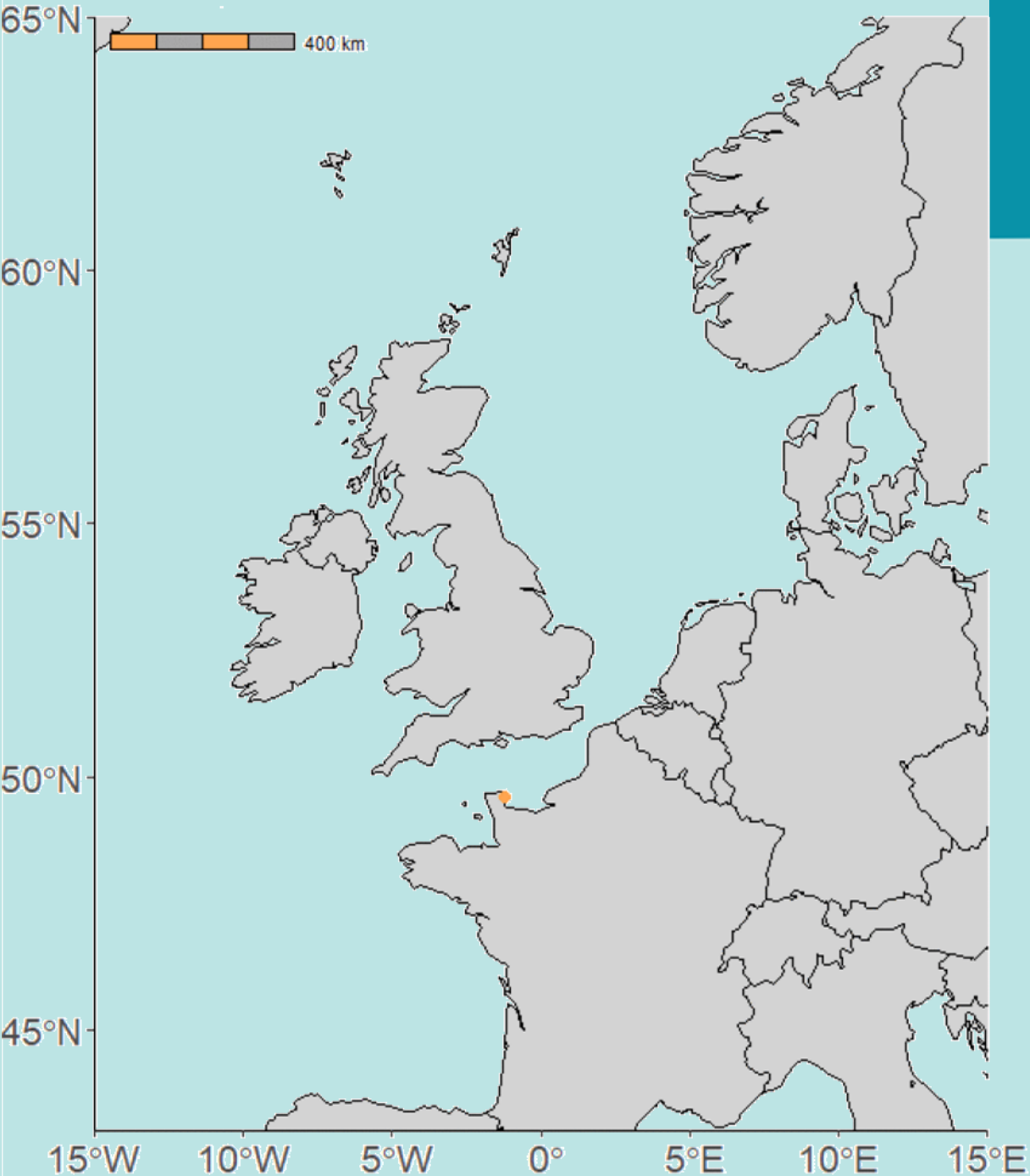
981  
records



1,108  
records

25,758 records total

Year: 1965



# The published literature

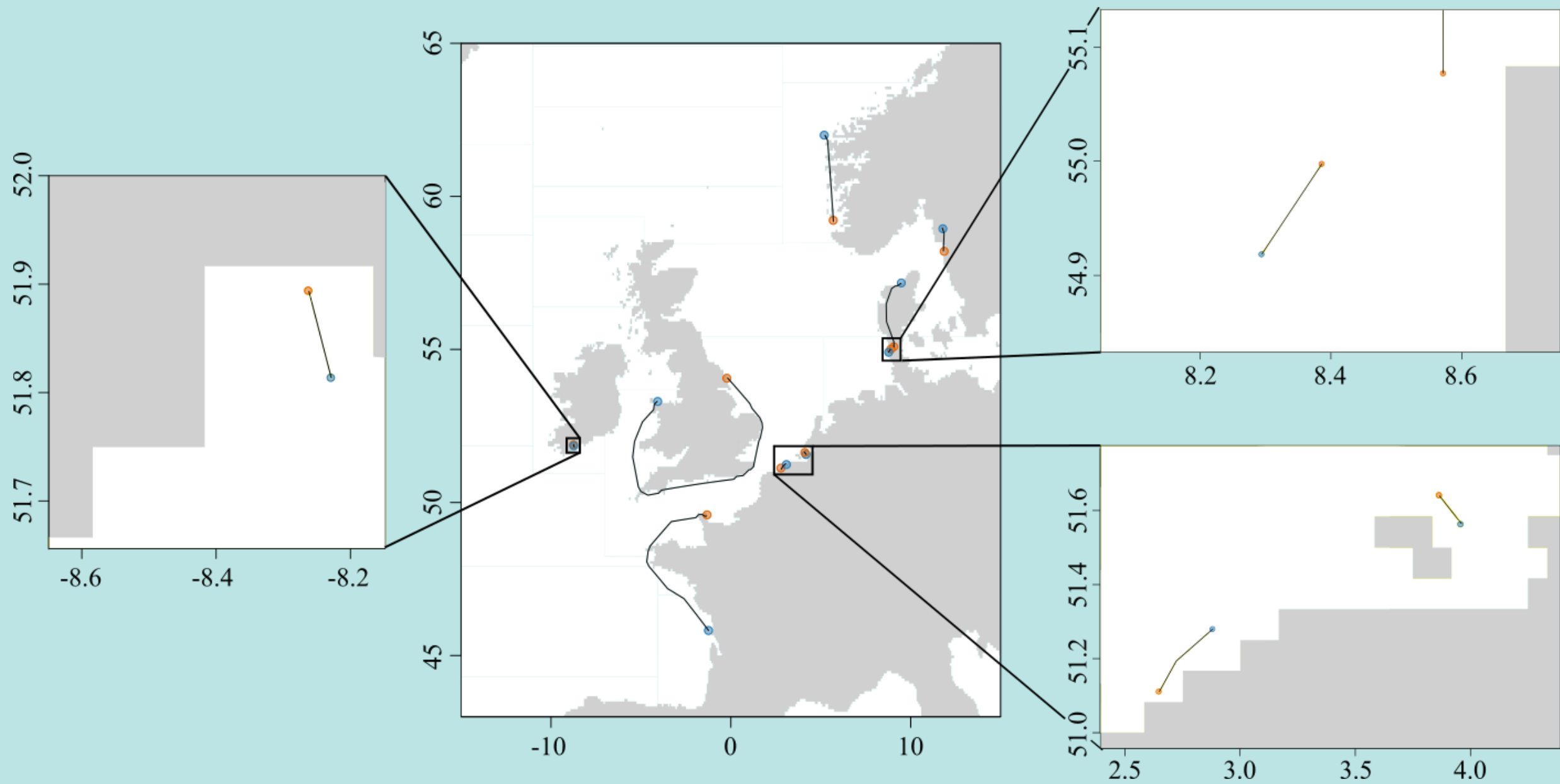
Published, peer-reviewed scientific literature searched for year and location of first record

Approximate coordinates taken from Google Earth Pro



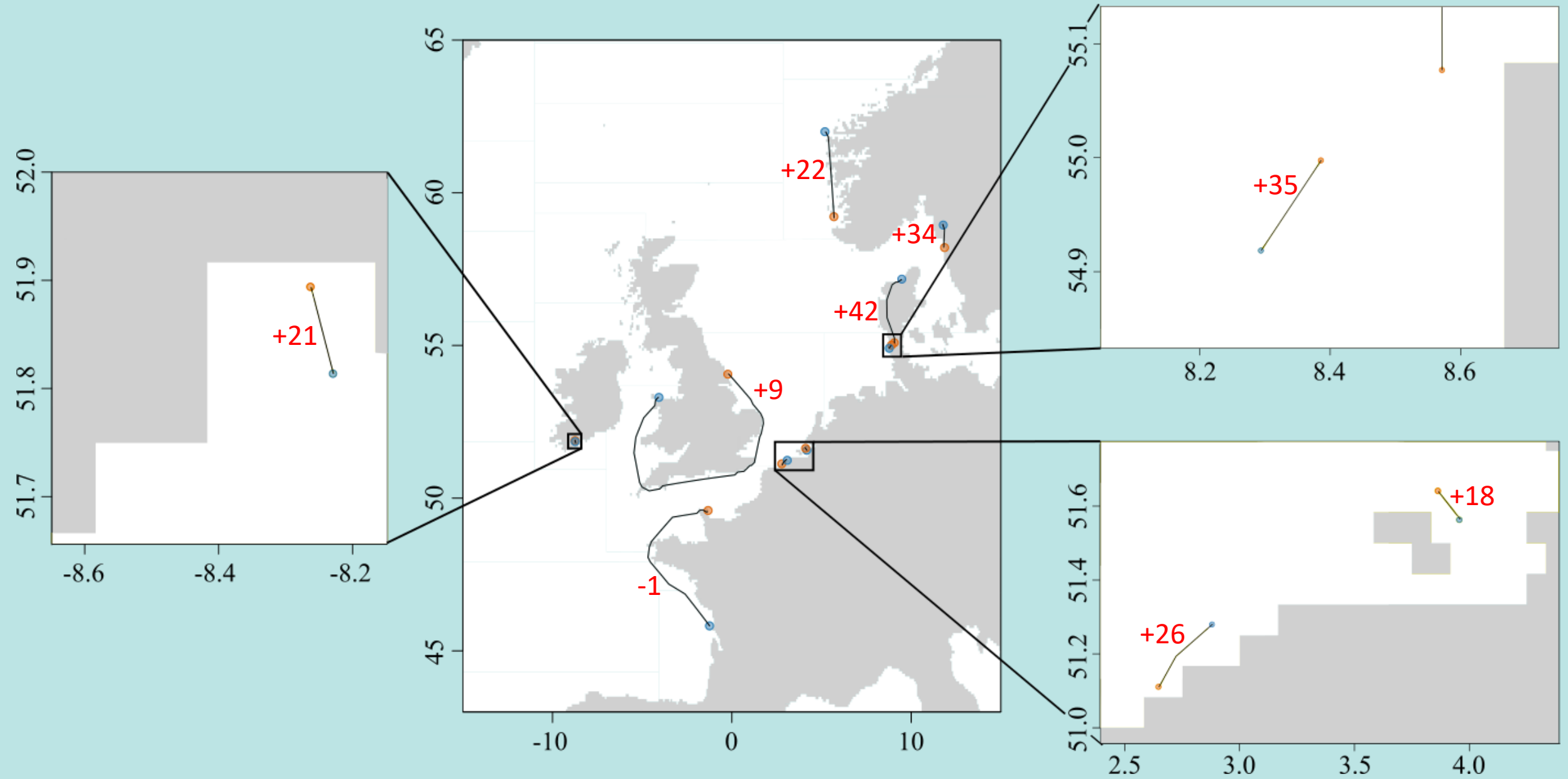


# (1) Identify discrepancies between data types





# (1) Identify discrepancies between data types



# Calculating spread

## Maximum rate of spread (Max-spread)

- Adapted from Mineur *et al.* (2010)
- Spatially explicit
- Calculates seaway distance of records from each year to the earliest record
- Distances filtered to only those representing an increase on the previous maximum distance, and linear regression applied

## Calculated twice

(i) Location of first record from the database

- Saint-Vaast-la-Hougue, France

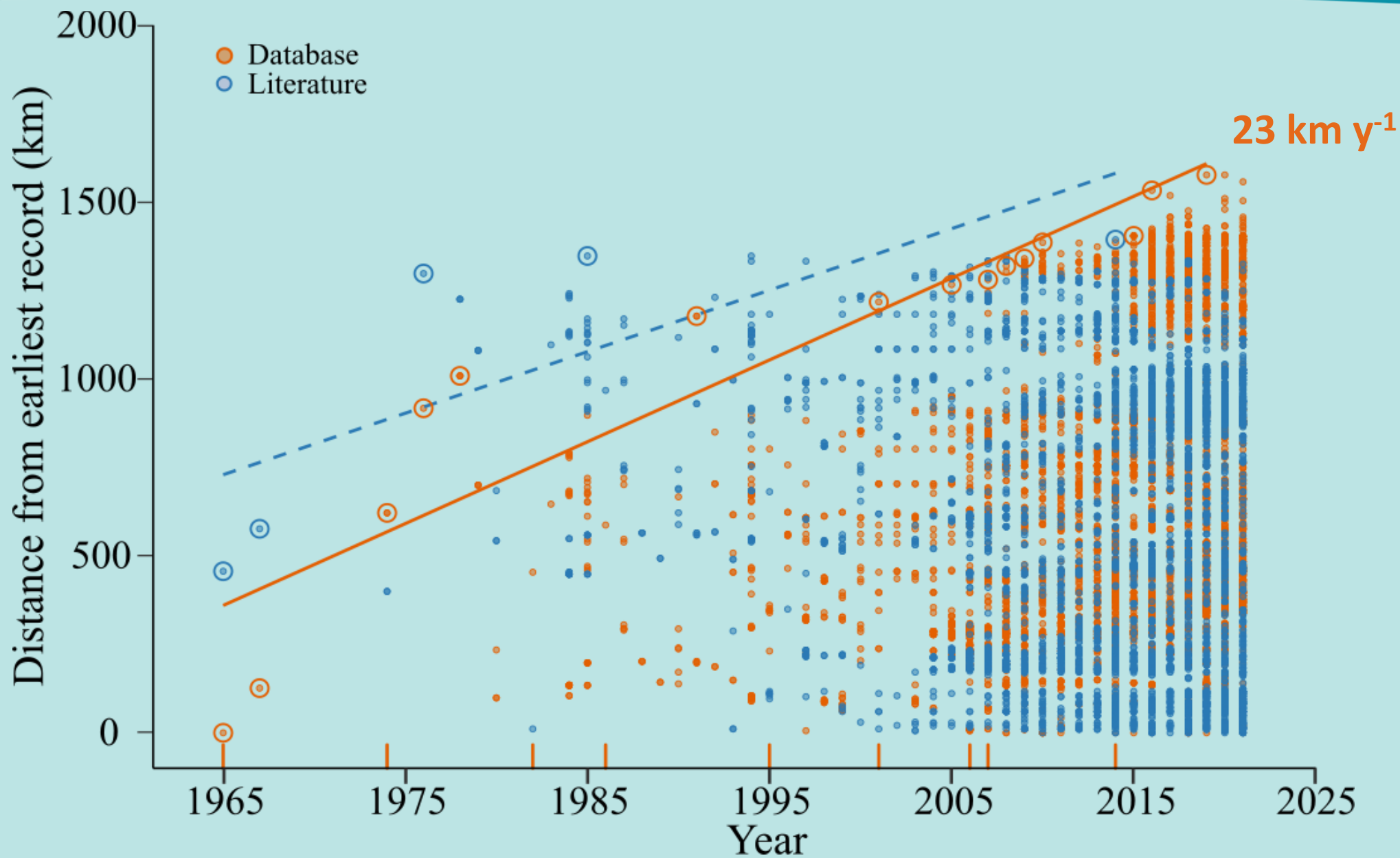


(ii) Location of first record from the peer-reviewed literature

- Oosterschelde Estuary, The Netherlands



## (2) Do discrepancies affect spread estimates?



# Implications for data

## Despite disparities in data, spread estimates were not significantly different

- High use of databases in invasive and non-native science
- > 1,750 peer-reviewed publications using GBIF data as of 2016

## *M. gigas* is a conspicuous species

- Do spread estimates remain the same for less conspicuous species?



## ‘Cleaning’ and systematic curation of records is a valuable exercise

- Comparison provides an opportunity to critique accuracy and improve quality
- Improve understanding of invasion dynamics

# Thank you for your attention



@charley\_clubley



charlotte.clubley@plymouth.ac.uk

**Acknowledgements:** PlyMSEF  
(Plymouth Marine Science and  
Education Foundation)



**UNIVERSITY OF  
PLYMOUTH**



Centre for Environment  
Fisheries & Aquaculture  
Science